DOCUMENT RESUME

ED 212 487 SE 036 234

AUTHOR	Panchyshyn, Robert; Enright, Brian
TITLE	An Investigation of Word Frequency in Mathematical Word Problems in Basal Mathematics Textbooks, Grades One Through Eight.
INSTITUTION	Western Kentucky Univ., Bowling Green.
PUB DATE	Nov 81
NOTE	22p.; Paper presented at the Meeting of the Mid-South Educational Research Association (Lexington, KY, November 12, 1981). Contains broken type in references.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Educational Research; Elementary Secondary Education; *Mathematical Vocabulary; *Mathematics Education; *Mathematics Instruction; *Problem Solving; *Textbook Research; Textbooks
IDENTIFIERS	*Mathematics Education Research; Word Counts; Word Familiarity; *Word Problems

ABSTRACT

This research project was initiated to examine the vocabulary load contained in word problems appearing in basal mathematics textbooks through a study of word frequency. Five leading basal mathematics series were used. Every word, phrase or sentence that resulted in computation was included. A total of 476,674 words were identified. Information with regard to the total number of words, the number of different words, and the extremely high percentage of words appearing one time per book is presented. (Author)

An Investigation of Word Frequency in Mathematical Word Problems in Basal Mathematics Textbooks, Grades One through Eight

Prepared by:

Robert Panchyshyn
Brian Enright

Department of Teacher Education
Western Kentucky University

Prepared for:

Mid-South Educational Research Association

Lexington, Kentucky

November 12, 1981

Funding for this project was partially provided by a Western Kentucky University Faculty Research Grant

Abstract

This research project was initiated to examine the vocabulary load contained in word problems appearing in basal mathematics textbooks through a study of word frequency. Five leading basal mathematic series were used. Every word, phrase or sentence that resulted in computation was included. A total of 476,674 words were identified. Information with regard to the total number of words, the number of different words, and the extremely high percentage of words appearing one time in the book is presented.

An Investigation of Word Frequency in Mathematical Word Problems in Basal Mathematics Textbooks, Grades One through Eight

INTRODUCTION

The problem undertaken in this study focused upon the word load of story problems contained in basal mathematics textbooks. The need for this determination stemmed from concern expressed by educators about the lack of success of children in word (story) problem solution. This issue has also become one of the central issues of the National Council of Teachers of Mathematics in their agenda for action in the 80's. Questions relative to vocabulary, word perception, and word frequency have been examined.

Identification of the Problem:

Mathematics teachers, whether in departmentalized teaching situations or in self contained classrooms continue to show great concern regarding the word problem solution achievement of their students. Results from the California Test of Basic Skills in Kentucky and other states indicate that a significant number of students who have developed adequate computational abilities consistently score lower on word problem tests.

Leading educational journals repeatedly cite the word problem dilemma, and a significant number of mathematics and reading experts have and continue to spend research and writing time discussing and proposing solutions (Repp, 1960, Thomas and Robinson, 1977, Noddings, 1980). Purusal of the literature by the current investigators has

revealed that the prevailing theme regarding word problem solution reflects a necessity for mathematics teachers to recognize the conciseness, abstractness and complex relationships that characterize mathematical material, then make special provisions for instruction. Any effort to accomplish this is commendable and that this is being attempted is not questioned by these researchers. The issue addressed in this project is basic word perception, its relationship to successful word problem processing and the ensuing solution.

The main emphasis of this research project has centered on the problem of basic word recognition developed from the theme of word perception and instructional level. The researchers theorize that mathematical writing contains an overwhelming burden of words causing the "directional attention" of the reader to be diverted. Directional attention is interpreted as being the ability to focus ones' cognitive processes on a specific task and toward a clearly defined outcome. The researchers feel strongly that directional attention of the readers in question should be focused on the solution of the word problem rather than on the word burden contained in their textbooks.

Specifically this research project was initiated to examine the vocabulary load contained in word problems appearing in basal mathematics textbooks through a study of the frequency of words used. Every word, phrase or sentence that resulted in computation was included in the word count and subsequent evaluation.

REVIEW OF LITERATURE

The primary consideration of this study was to evaluate,

through a frequency examination the word load in five basal mathematics textbook series, grades one through eight. As indicated earlier, a substantial amount of information reflecting the necessity for the development of a systematic approach to word problem solution is available for study (Thomas and Robinson, 1977, Noddings, 1980, Munro, 1977, Hiesch and Lewinger, 1975, Dunlap and McKnight, 1978). Review of the literature presented in this section will deal with the issue of fluency in word recognition.

That children learn to read through the application of a wide variety of materials and methodologies cannot be questioned. It is the processes of basic perception and the subsequent transferal through some comprehension process culminating in problem solution which is being scrutinized. It is apparent that the problem solver, prior to computation, and prior to application of mathematical reasoning reflecting the digestion of the technical and sub-technical vocabulary in the passage, must recognize all the words in the passage.

Dr. William Gray in his historic model of reading identifies word perception as the basis, the key to success in reading. The subsequent skills of comprehension, reaction and fusion are all dependent on success in this most basic skills area (Clymer, 1968). Word perception includes recognition, correct pronunciation, and meaning intended by the author. Fluency in perception is undoubtedly one of the key factors leading to successful management of written passages by a reader.

The publishers of basic reading material obviously support this philosophy in the development of basic readers. This is

reflected by the great emphasis placed on the controlled vocabulary aspects of the materials they publish (Hall, Ribovitch, and Ramig, 1979). Further evidence regarding the necessity of a high rate of instant recognition as a requisite for successful developmental reading and the acquisition of related skills lies in the generally accepted criterion for "instructional reading level." Most professionals in the reading area agree that a child placed in his best instructional level should recognize 95% of the words in a selection without aid from anybody and without the apparent use of any word attack skills (Harris and Sipay, 1980, Woods and Moe, 1977, Lapp and Flood, 1978). The word load remaining which includes new or unknown words is only five percent, a mere one word out of twenty that needs some special attention prior to engagement with it in running text. In addition, it is often desirable for the "new" or "unknown" word to contain elements of word skill techniques known or being currently taught and developed, and/or to include words that have a high frequency of use which can be introduced through some type of "look-say" process (Harris and Sipay, 1980).

Singer (1976) indicates that the reader, in order to overcome the barrier presented by words in the form of printed symbols develops "word-recognition-substrata." As he develops reading skill his ability to acquire facts to form concepts and generalizations continuously increases. He uses his "reasoning-in-context substrata" to process the printed information. Singer goes on to state:

> In a modern instructional program, word-recognition substrata, word-meaning-substrata, and resoning-in-context substrata are developed simultaneously. But

> as a result of mental capability, training and practice, the time and mental energy required for word-recognition-substrata in the process of reading gradually decrease while the time and energy for employed for word-meaning-substrata and reasoning-in-context substrata gradually increase. (p. 314)

If the condition described above is accurate, then any inability or weakness in word-recognition tasks generates an interference with meaning and contextual development needed. In problem solving the individual is called on to conceptualize all the information, verbal and quantitative, in an entire passage and integrate the given information for problem solution. In order to have success he must simultaneously employ a number of different reading strategies. Singer (1976) states that an interference in comprehension and continuity of thought does occur when the reader expands an excessive concentration of energy in overcoming word-barriers:

> However, when a strange or unfamiliar word does appear, an individual has to mark time in his mental processing of the content and switch most of his mental energy to the operation of his word-recognition-substrata. (p. 312)

The phenomenon described above becomes much more serious in word problem format than in the usual kind of passages encountered. As previously stated, word-meaning-substrata and reasoning-in-context substrata must be taking place to enable the reader to accumulate, classify and store the information, then process it all into mathematical conceptualization for computation. If an undue burden of once used words exists an unnecessary interference is created. This burden could be avoided by the careful inclusion of a substantial number of familiar words.

Goodman (1976) defines reading as a "psycholinguistic guess-

ing game" explaining that:

> Reading is a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the readers expectation. (p. 498)

However, it must be noted that the process described above can take place only upon the development of a level of proficiency which allows the reader to process input of graphemic symbols or sequences simultaneously with an adequate number of available cues from within oneself or from the material at hand. This kind of processing is highly dependent on a sufficient sight vocabulary or vocabulary redundancy which enables the reader to attend to the task at hand.

Although Goodman indicates that highly skillful readers require less precision, he does state"... more accurate first guesses (are) based on better sampling techniques, greater control over language structure, broadened experiences and increased conceptual development" (p.504). In word problem reading the child must actively recognize, swiftly and accurately develop, collect and sort word concepts all to be intercorrelated with necessary mathematical thinking. Psycholinguistic guessing in this kind of reading must be replaced with recognition precision coupled with sufficient built-in ease of discourse which allows maximum attention focused on mathematical conceptualization.

Successful application of known reading strategies and techniques for mathematical problem solution must begin and be developed with minimal attention given to basic word perception and maximum attention directed toward problem conceptualization. The factor of attention in comprehension is cited by Mathewson (1976) stating,

"Attention is by definition prior to comprehension. It must be secured before reading can take place" (p. 660). Again this illustrates that if an excessive number of unknown words appear in a passage, directional attention could be shifted from problem processing to basic word perception.

Additional evidence to corroborate the thesis presented comes from the continued interest in high frequency word list development. More than one hundred such word lists have been constructed to consider a wide variety of questions. Basically most of these lists are used to evaluate word load in reading materials, readability, and/or for development of reading materials that help teach these words in context and in isolation. Usually a high percentage of the running words in given passages are those appearing on the various word lists.

The Dolch list developed in 1932 was designed to give teachers a short list of often used words which comprise a high percentage of running words in a passage. In a replication study Johns, Edmond, and Mavrogenes (1977) revealed that in first through sixth grade level basal readers, Dolch words accounted for seventy to fifty-nine percent of words used. With a modification of selection, Johns' words ranged from sixty-six to fifty-five percent. Guzak (1977) also reported similar agreement.

Several studies have reexamined the original list by Dolch in an effort to up-date it and compare it to comtemporary writing. Johnson (1971) using the Kucera-Francis Corpus made up of words to which the adult reader is exposed made such comparison. It reveals that only thirty-seven percent of the words appearing on

the Dolch list are not on the Kucera-Francis list indicating that comparative studies of passages using either list would probably be in order. This would be especially true of the first forty-seven words which are identical on the two lists.

In another study Johns (1976) compiled a word list based on four different important lists including American Heritage Intermediate Corpus, Kucera-Francis Corpus, Murphy, and Durr. Of the two hundred twenty-six words which were common to at least three of the word lists only thirty-five were not on the original Dolch word list.

The presentation in this section was developed primarily to reveal the consistancy existing throughout the various lists commonly being used. Initial data and comparisons will be made with the more historic Dolch list, however future comparisons will include other contemporary word lists.

This literature review has been developed to identify important considerations relative to the concern of mathematical word problem solution. The importance of developing systematic approaches to information processing, acquisition of technical and sub-technical vocabulary and any other supportive techniques has not been questioned. The objective was to demonstrate the importance of basic word perception and its relation to problem processing. Finally word frequency information was included to justify the basis for frequency study and comparison.

METHODOLOGY

The major purpose of this research was to develop a corpus

of words which, based upon vocabulary used in five basal math series could constitute the mathematics vocabulary necessary for word problem (story problem) solution. Secondly, this corpus of words was to be analyzed to determine the frequency of each word on the derived list. Special attention was given to those words which appear infrequently.

The use of computers to count and sort words has been functional for over two decades. This technological aid increases the numbers of words that can be reviewed at one time, as well as reduces the probability of error occuring through the laborius task of manually counting words (Moe, 1980). The computer used in this study is an IBM-370.

Two different programs were utilized to insure the accuracy of the findings. The first program is entitled "TUIT" and was developed by Dr. Beverly Madron while at Western Kentucky University (Madron, 1979). This program provides a variety of information regarding selected reading passages, however, only the component for sorting and counting of words has been utilized at the present time. Secondly, the Statistical Analysis System (1979) was used to verify the "TUIT" program findings. The second program provided additional data on cummulative totals and percentages. Both programs operate in the batch mode.

DATA COLLECTION

Five basal mathematics series (Grades 1 to 8) were selected for inclusion in the study. This represented half of the series appearing on the current Kentucky State Adoption List, It was

felt this number of texts would provide an adequate sampling of all texts. The five series included represented some of the most commonly used mathematics material throughout Kentucky and the entire country. At the time of this report three books of the forty possible books were not included in the findings. These books were not available at the time of the analysis, but will be added at a later date. However, it was felt that these books would not significantly change the overall findings, nor was it felt that they would be significantly different from the other books in their respective analysis combinations. The books excluded at the present time are Houghton-Mifflin, grades 1 and 4 and Silver, Burdett, grade 8.

Each text was reviewed by the investigators and every word, phrase, or sentence that resulted in a computation was underscored. This was done to eliminate those words that were used to teach or develop computation skills. Each text was then forwarded to the Computer and Information Services Office at Western Kentucky University where all the underscored words, phrases, and sentences were keypunched on standard computer cards. A grand total of 476,674 words were used. The words were coded by name of the series and by grade level.

DATA ANALYSIS

The data set was analyzed using both the "TUIT" and the "SAS" programs. The data was sorted according to series by grade levels; series across grade level; and grade level across series. Lists compiled for each dimension provided data with regard to frequency

and the relationships of each word to the total. The word list for each analysis was produced separately, but is not included in this report due to the size of said lists.

RESULTS

The results of this study appear in five separate tables. Initially, the investigators were interested in differences between series. Results in Table 1 indicate that there does appear to be substantial differences in the word load between series (*note missing books mentioned earlier).

Table 1: Comparison by Publisher of Total Number of Words and Unique Words.

Publisher	Total Number of Words	Number of Unique-Words
Holt, Rinehart, & Winston	126,244	5,722
*Houghton-Mifflin	70,079	4,513
Macmillan	77,769	3,698
Scott, Foresman	105,423	5,733
*Silver, Burdett	97,159	5,422

Interpolating for the series with missing books leaves only the Macmillan series with a significantly smaller word load, both in total word load and in the number of unique words (ie. individual words that appear one or more times).

Table 2: Comparison by Grade Level of Total Number of Words and Unique Words (Across Publishers).

Grade Level	Total Number of Words	Number of Unique-Words
*1st	4,717	357
2nd	11,829	921
3rd	47,562	2,909
*4th	57,214	3,709
5th	82,276	4,774
6th	93,743	5,414
7th	105,328	5,760
*8th	74,023	4,775

Table 2 presents the findings by grade level across series. Again interpolating for the missing books, a steady growth of both the total number of words and the number of unique words is indicated. This growth is quite marked at the primary grade levels showing over a seven thousand total word increase between first and second grade, and an increase of approximately thirty six thousand words between second and third. The totals then somewhat level off in grades five to eight. This is especially significant in that during the period of primary development of basic word perception skills, the eronmous jump in word load does put significant pressure on the student.

Table 3: Percent of Unique Words Appearing An Average of One (1) Time Per Book By Publisher Across Grade Level.

Publisher	Number of Unique Words	Number of Words Appearing Once	% of Words Appearing Once
Holt, Rinehart, & Winston	5,722	4,428	.77
*Houghton-Mifflin	4,513	3,564	.79
Macmillan	3,698	2,808	.76
Scott, Foresman	5,733	4,608	.80
*Silver Burdett	5,422	4,356	.80

Table 3 provides data regarding the internal structures of the corpus of words by series. Two points become immediately clear. First, the number of words appearing an average of one time per grade with each series is fairly consistent, with Macmillan again having the lowest total and the lowest percent (76%). Secondly, and possibly more important is the high percentage of words appearing once per book (from 76% to 80%). This high figure greatly increases the possibility that more than five percent of the words will be unknown and therefore, as mentioned earlier in this paper, the difficulty level of the material will increase substantially, resulting in errant directional attention of the student.

Table 4: Percent of Unique Words That Appeared an Average of One (1) Time Per Book at Grade - Levels Across Publishers.

Grade Level	Number of Unique Words	Number of Words Appearing Once	% of Unique Words
*1st	357	252	.71
2nd	921	684	.74
3rd	2,909	2,124	.73
*4th	3,709	2,736	.74
5th	4,774	3,456	.72
6th	5,414	4,032	.74
7th	5,760	4,176	.73
*8th	4,775	3,528	.74

Table 4 provides data comparing the number of unique words with the number of words having only a single appearance. The findings demonstrate a relatively stable percentage (70%+) of unique words appearing in each successive set of materials indicating that even when the total number of words increased substantially, there was little effect on unique word percentage. Again the set of data demonstrates that the word burden increased substantially between first and second grade and dramatically between second and third, and further that the reader dependence on redundant or high frequency words must be negatively affected by the minimal percentage change and the continued high rate of unique word appearance.

Table 5: Comparison of Words Appearing an Average of 1 Time Per Book or More with the Dolch List by Grade Level Across Publishers.

Grade	Number of Unique Words	Number of Words Appearing Once	% of Words Appearing Once From Dolch
*1st	357	252	21%
2nd	921	684	8%
3rd	2,909	2,124	1%
*4th	3,709	2,736	1%
5th	4,774	3,456	1%
6th	5,414	4,032	1%
7th	5,760	4,176	1%
*8th	4,775	2,340	1%

Table 5 provides the results of a comparison of once used words with the Dolch list. As noted earlier the Dolch list represents words most frequently encountered by elementary age students. It becomes clear that the Dolch words make up very few of the words needed to solve the word problems.

DISCUSSION

The purpose of this study was to identify a corpus of words needed to solve mathematics word problems as they appeared in five leading basal series. The investigators also studied the composition of the lists in an attempt to determine if the word selection by authors of basal series interfered with student success of

solving word problems. The investigators have earlier in this paper presented the concept that instructional readability of material may be directly affected by the percent of words that are not instantly recognized. Therefore, the investigators contend that if a disproportionate percent of words were not immediately recognizable to the student, the instructional readability level of the material becomes more difficult. The investigators further contend that if words were not immediately perceived by the student, the student's "directional attention" would be diverted from processing for computation of the word problem toward the recognition of the words in the passage.

The results of this study indicate that a significantly disproportionate number of words appear one time per grade level across series. Whether one examines by grade level or by series, a consistant finding is that over 70% of the unique words in any of the specific corpus of words appear only one time. Further analysis indicates that a majority of these words do not appear on the Dolch high frequency list, thereby reducing the probability of instant sight recognition.

It is reasonable to assume that words which appear frequently are more apt to become sight words and consequently useful aids in more effieient and successful processing word problems. It is also reasonable to assume that words used infrequently have less chance of becoming a part of a student's sight vocabulary and consequently could become obstacles in word problem solution. Given this set of assumptions one might choose to consciously design mathematical word problems using words of high frequency and/or using words frequently wherever possible.

FUTURE CONSIDERATIONS

Much additional investigation is needed in this area. A study which would consider the phonetic structure of the words used in basal mathematics series might indicate whether the student has yet mastered the phonetic skills necessary for independent analysis.

A study which would classify high frequency words into catagories such as nouns, verbs, adjectives, and adverbs to develop a functional system for production of materials using more high frequency words and repeating words more often may be beneficial.

A carefully designed study using rewritten current word problems using concepts, settings and vocabulary already known by the child would be useful. This study would help determine the amount of interference generated by word problems written in the current fashion.

Much attention needs to be paid to the difference between the successful reader and the unsuccessful one with regard to the solution of mathematic word problems. It may not be functional to consider the same approach or for that matter the same series for each group. The unsuccessful reader of word problems may need a substitute program more highly structured and more closely controlled for vocabulary.

Finally, it should be noted that the findings presented in this report are only preliminary. Subsequent evaluation of the data collected should reveal a number of important insights into the development of mathematics materials and methodologies. The research team intends to continue evaluation of the data and present additional information as it becomes available.

References

Bolster, L.C., et.al. Mathematics Around Us. Glenview, IL: Scott, Foresman and Company, 1978.

Clymer, J. What is reading: some current concepts. In H.M. Robinson (Ed.), Innovations and change in reading instruction. Chicago: The National Society for the Study of Education, 1968.

Duncan, E.R., et.al. Mathematics. Boston: Houghton Mifflin Company, 1978.

Dunlap, W.P. & McKnight, M.B. Vocabulary translations for conceptualizing math word problems. Reading Teacher, 1978, 32, 183-189.

Forbes, J.E. Thoburn, T., etal, Macmillan Mathematics. New York: Macmillan Publishing Co., 1978.

Goodman, K.S. Reading a psycholinguistic guessing game. In S. Singer & R.B. Ruddell (Ed.), Theoretical models and processes in reading. Newark, N.J.: International Reading Association, 1976.

Guzak, F. J. Diagnostic reading instruction in the elementary school. New York: Harper and Row, 1972.

Hall, M.A., Ribovich, J.K & Ramig, C.J. Reading and the elementary school child. New York: D. Van Nostrand, 1979.

Harris, A.J. & Sipay, E.R. How to increase reading ability. New York: Longman, 1980.

Hirsch, R. and Lewinger, M. Words to grow on. Teacher, 1975, 96, 6, 72-77.

Johns, J.L. Updating the Dolch basic sight vocabulary. Reading Horizons, 1976, 16, 104-111.

Johns, J.L., Edmond, R.M. & Maurogenes, N.A. The Dolch basic sight vocabulary: A replication and validation study. Elementary School Journal, 1977, 78, 1, 31-37.

Johnson, D.D. The Dolch list reexamined. Reading Teacher, 1977, 24, 5, 449-457.

Lopp, D. & Flood, J. Teaching reading to every child. New York: Macmillian, 1978.

Madron, B. TUIT. In Design of a computer base system for research in and teaching of literature. George Peabody College for Teachers, Unpublished Dissertation, 1979.

Mathewson, G.C. The function of attitudes in the reading process. In H. Singer & R.B. Ruddell (Ed.), Theoretical models and processes of reading. Newark, N.J.: International Reading Association, 1976.

Moe, A.J. Analyzing text with computers. Educational Technology, 1980, 20, 7, 29-31.

Munro, J. Language abilities and maths performance. Reading Teacher, 1977, 32, 8, 902-915.

Nichols, E.D., et.al. Holt School Mathematics. New York: Holt, Rinehart and Winston, PUblishers, 1978.

Noddings, N. Word problems made painless. Creative computing, 1980, 6, 9, 108-113.

Repp, F. C. The vocabularies of five recent third grade arithmetic textbooks. Arithmetic Teacher, 1960, 7, 7, 128-132.

Singer, H. Conceptualization in learning to read. In H. Singer & R.B. Ruddell (Ed.), Theoretical models and processes of reading. Newark, N.J.: International Reading Association, 1976.

Statistical Analysis System Institute Inc., Raleigh, N.C. (release 79.5)

Thomas, E.L. & Robinson, H.A. Improving reading in every class. Boxton: Allyn and Bacon, 1977.

Vogeli, B.R., Leblanc, J.F., et.al. Mathematics for Mastery. Glenview, IL: Silver Burdett Company, 1978.

Woods, M.L. and Moe, A.J. Analytical reading inventory. Columbus: Charles E. Merrill, 1977.

WESTERN KENTUCKY UNIVERSITY
The cost of printing this publication by Western Kentucky University was paid from state funds KRS 57 375.